



NRL/FR/5520--96-9805

IVOX—The Interactive VOice eXchange Application

JOSEPH P. MACKER

*Communication Systems Branch
Information Technology Division*

R. BRIAN ADAMSON

*NEWLINK Global Engineering Corporation
6186-B Old Franconia Road
Alexandria, VA*

February 20, 1996

19960314 087

Approval for public release; distribution unlimited.

DTIC QUALITY INSPECTED 1

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave Blank)		2. REPORT DATE February 20, 1996		3. REPORT TYPE AND DATES COVERED	
4. TITLE AND SUBTITLE IVOX—The Interactive VOice eXchange Application				5. FUNDING NUMBERS PE - 63792N TA - R1889	
6. AUTHOR(S) Joseph P. Macker and R. Brian Adamson					
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Research Laboratory Washington, DC 20375-5320				8. PERFORMING ORGANIZATION REPORT NUMBER NRL/FR/5520-96-9805	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Office of Naval Research 800 North Quincy Street Arlington, VA 22217-5660				10. SPONSORING/MONITORING AGENCY REPORT NUMBER	
11. SUPPLEMENTARY NOTES *NEWLINK Global Engineering Corporation, 6186-B Old Franconia Road, Alexandria, VA 22310					
12a. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited.				12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words) A flexible, low data rate network voice terminal application and a set of new network integration techniques have been designed and developed at NRL for the support of real-time interactive voice communication over distributed, computer data networks. The NRL Interactive Voice Exchange (IVOX) uses advanced voice compression techniques to maintain a low data rate throughput requirement. The low data rate feature of IVOX also allows for the use of voice communication over existing computer networks without a significant impact on the other data communications (e.g., e-mail, file transfer). IVOX provides a simple graphical user interface for call setup and management. IVOX allows for cross computer platform interoperability with versions for Sun SPARCStation, Silicon Graphics, and Hewlett Packard workstation. Since its conception and development, IVOX has become an operational fleet software component and has played a key role in numerous research projects and demonstrations. IVOX was adopted as a core feature of the Joint Deployed Intelligence Support System (JDISS). During 1995 Joint Warrior Interoperability Demonstration (JWID '95), IVOX successfully demonstrated integrated network voice capability between Tactical Aircraft Mission Planning System (TAMPS) and Common Operation Mission Planning and Support Strategy (COMPASS) workstation terminals. Additionally, IVOX was demonstrated from an operational NRL networking booth during the 1995 Armed Forces Communications and Electronics Association (AFCEA 95) convention. IVOX enhancements have been integrated into several research projects including the NRL Data and Voice Integration Advanced Technology Demonstration (DVI ATD) and the Common System NATO Interoperability (CSNI) project.					
14. SUBJECT TERMS Packet voice Internetworking Integrated networking Voice systems Low rate vocoding				15. NUMBER OF PAGES 17	
				16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT UNCLASSIFIED	18. SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED	19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED	20. LIMITATION OF ABSTRACT UL		

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std Z39-18
298-102

CONTENTS

INTRODUCTION.....	1
BACKGROUND	1
DESIGN	2
Overview of Operation.....	2
Network Voice Delivery Requirements.....	3
Features.....	4
Voice Algorithm Issues.....	6
Call Setup and Management Protocol	8
Voice Packet Buffering and Reordering.....	11
SUMMARY	12
Accomplishments	12
Future Directions.....	12
ACKNOWLEDGMENTS	12
REFERENCES.....	13

IVOX—THE INTERACTIVE VOICE EXCHANGE APPLICATION

INTRODUCTION

The design and development of the Interactive VOice eXchange (IVOX) computer software application is described in this report. IVOX provides real-time interactive voice communication over computer data networks by using advanced voice compression techniques to maintain very low data rate throughput requirements. The low data rate feature permits effective voice communication over existing computer network connections without a significant impact on other data communications (e.g., email, file transfer). IVOX provides a simple graphical user interface for call setup and management. It also allows for cross computer platform interoperability with versions for Sun SPARCStation, Silicon Graphics, and Hewlett Packard workstations. Additional support for Digital Equipment Corporation (DEC), Apple Macintosh, and Microsoft Windows platforms is in development [1].

BACKGROUND

Digital voice communication systems have typically required dedicated allocation of communication resources that are separate from those used for other data communication applications, such as tactical data exchange and electronic file transfer. Often, those dedicated resources are significantly underutilized during periods of low demand for voice communication. Open systems computer internetwork technology with protocol stacks such as the Transmission Control Protocol/ Internet Protocol (TCP/IP) suite has made it possible for a large number of users to efficiently and dynamically share heterogeneous communication media and networks [2]. This technology presents practical, flexible resource sharing and utilization capabilities, and is presently being applied to provide a general purpose data communication services within the Navy communication infrastructure [3]. It is possible to support useful voice communication over computer data networks given sufficient data throughput capacity, acceptable data delivery latency at the link layer, and proper application design [4]. Integration of voice communication services along with other data services offers the following advantages.

- More efficient use of available limited communications bandwidth and resources.
- Reduced need for multiple communications management systems.
- Potential for integrated multimedia computer applications to support distributed mission planning and execution. (e.g., distributed interactive conferencing and presentation systems).
- Lower communication resource cost by reducing the need for dedicated, leased voice coordination circuits (e.g., INMARSAT channels).

The NATO Communication Systems Network Interoperability (CSNI) [5] and the NRL Data and Voice Integration Advanced Technology Demonstration (DVI ATD) [6, 7] projects developed and demonstrated network architectures that support integrated digital communication services, including voice, with application of open systems network protocols and technology. The network connectivity provided by these research projects consists of low data rate, radio frequency (RF) communication links. Both projects used connectionless, packet switched routing and data delivery protocol frameworks. CSNI used the International Standard Organization (ISO) Connectionless Network Protocol (CLNP) stack, and the NRL DVI ATD used the TCP/IP protocol stack.

Some experimental and commercially available computer applications have been developed to provide voice communication over connectionless computer networks (e.g., InPerson, ShowMe, Visual Audio Tool (vat), Network Voice Terminal (nevot)). However, the use of relatively high data rate voice coding within these existing network voice applications places unnecessary demands on performance and limits participation for mobile users and network sites with relatively low bandwidth resources such as found in many tactical communication systems. Applying adaptive, low rate compression algorithms to network voice communications is an enabling technology for users with low bandwidth resources and enhances the performance of higher bandwidth systems under loaded conditions. This is the fundamental principle followed in the development of IVOX; use of advanced low data rate voice compression and use of open systems computer network protocols.

DESIGN

The design of IVOX was conducted with the features and limitations of network data services in mind. In particular, the additional limitations imposed by low data rate tactical connectivity were given special consideration. The following is a list of design goals for IVOX.

- Use existing computer platforms without modification
- Provide a simple graphical user interface with familiar telephone call paradigm
- Robust and efficient call setup and management protocol
- Provide multiple data rate vocoder algorithms for multiple levels of voice quality and capability of operation over a wide range of network connections
- Modular software design for easy integration of additional vocoders and other features
- Multiple modes of operation, e.g., half-duplex/ full-duplex, real-time/ non-real-time
- Flexible set of user controllable parameters for evaluation over different data networks.

Overview of Operation

IVOX digitizes voice by using the built-in computer audio hardware and then, using specialized speech encoding algorithms, compresses the audio data to continuous data rates as low as 600 bits-per-second (bps). Additionally, IVOX uses a silence detection technique to reduce the average data rate to even lower rates. For example, IVOX uses the Federal Standard 1015 Linear Predictive Coding (LPC) speech compression to achieve a data rate of 2400 bps. With silence detection removing the gaps between words or during pauses in speech, the typical measured data rate during *active* portions of conversational speech is reduced to approximately 1300 bps. Longer pauses and exchanges in conversation make the longer term average data rate much lower than this (Note: Some additional data capacity must be allocated to allow for network protocol overhead).

IVOX also supports operation with external voice compression hardware through a modular software interface [8]. This reduces the load on the workstation CPU during voice communication. The prototype hardware that has been currently demonstrated connects to the computer via the serial

port. It is envisioned that the hardware compression circuitry could be packaged as a plug-in bus card (e.g., EISA, PCI, PCMCIA) depending on the workstation requirement. Use of the voice compression hardware has allowed IVOX to operate on platforms with no built-in audio capability.

IVOX communicates over the Internet using the User Datagram Protocol (UDP) encapsulated in Internet Protocol (IP) packets. The connectionless UDP transport was chosen over the reliable, connection oriented TCP because TCP's reliability and flow control mechanisms (ACK-based, go-back-n retransmission) were prohibitive to real-time communication over low data rate links (Note: In the case of CSNI, CLNP's connectionless transport protocol (TP0) was chosen over the connection-oriented, TP4. The UDP/IP suite provides unreliable, unordered best effort delivery of data packets. An in-band signaling protocol is used to negotiate call setup and for subsequent communication session control. IVOX also provides a number of user controllable parameters for evaluating low data rate voice communication with connectionless datagram delivery. Example parameters include:

- Number of vocoder frames per packet
- Resequencing window time setting
- Half-duplex or full-duplex operation
- Real-time and nonreal-time interactive modes.

Network Voice Delivery Requirements

Some fundamental differences exist between the synchronous, dedicated communication channels typically used for digital voice communication and the data delivery service provided by connectionless network protocols. These include:

- | | |
|------------------------|--|
| • Data delivery: | Synchronous bit stream vs asynchronous packets |
| • Communication delay: | Deterministic vs nondeterministic delay |
| • Error handling: | Bit error rate vs packet drop rate |

Data Delivery

Generally, current digital voice communication systems provide continuous synchronous delivery of voice data across a dedicated communication channel. The dedicated communication channel is typically designed to provide a fixed amount of bandwidth, and vocoding algorithms have been designed to provide the best voice quality for bit rates supported within this bandwidth. In the connectionless networking environment where communication bandwidth is dynamically shared in an asynchronous fashion among distributed users and applications, adaptive rate voice coding techniques (e.g., silence detection) can improve bandwidth utilization dramatically. Adaptive rate voice coding can allow for high voice quality while maintaining a lower average network throughput requirement. The synchronous nature of many current voice encoding schemes has led to a widespread perception that voice communication requires "stream" oriented data communications when the information content of conversational voice is actually bursty in nature. This low data rate, bursty source can be serviced effectively by an asynchronous, connectionless network fabric. In IVOX, we have added an adaptive rate enhancement to existing DoD voice digitization algorithms and used this as the primary mode for network voice communication [9].

Communication Delay

Dedicated connection oriented communication channels can provide deterministic delay in the delivery of voice data. With connectionless datagram delivery, there can be a significant amount of

nondeterministic variance in the interarrival times of data packets. Furthermore, it is possible that data packets are received in a different order than they were transmitted. Current IP service provides "best effort" delivery where all data flows are given the same quality of service (QoS). Research is being conducted and initial test systems are in place for standardized IP routing and data delivery techniques that bound factors affecting QoS such as delay variance [10, 11]. Such vendor-independent resource management techniques will provide proactive internetwork bandwidth allocation among and between application data streams. Until techniques are widely implemented throughout the Internet, for many media, the "best effort" service model can continue to provide adequate service for even real-time applications such as digital voice communication. IVOX has been designed and tested using network architecture with and without QoS and resource reservation capability.

Error Handling

Previous digital voice communication systems integrate robust error handling within the speech coding algorithm because the voice terminal application has had direct access and control of the physical communication media. In contrast, the application layer within connectionless datagram networks maintains independence from the underlying communication media. In a global internetwork architecture, this independence allows applications to communicate peer-to-peer across multiple, heterogeneous media. Lower layer protocols at the transport, network, and link layers usually assume the major responsibility for any error handling. Datagrams in error are either automatically retransmitted or dropped depending upon the protocol set used. As a result, the application layer (i.e., the voice terminal in our case) does not usually incur bit errors in its communication data stream but will need to be able to handle undelivered packets and reconstruct packet ordering. IVOX is designed to appropriately handle and recover from dropped data packets and reorder received packets as necessary.

Features

Graphical User Interface

IVOX features a simple, easy-to-use graphical user interface for call setup and management. The main window of IVOX is illustrated in Fig. 1. The IVOX user interface roughly follows the paradigm of placing a normal telephone call. To place a call to a remote IVOX terminal, the user types in the name of the remote host (or dotted decimal IP address) and clicks on the "CALL" button. The remote user is notified of the pending call with a ringing sound, and then is given the option of accepting or rejecting the call. If the remote IVOX terminal is *busy*, the caller is notified. "Caller ID" is provided in the hostname display for incoming calls. Other telephone-like features such as "call waiting" and "voice mail" are planned for future versions of IVOX.

As an integral part of the user's workstation environment, IVOX potentially offers many features beyond that of a simple telephone service. These include voice messaging integrated with other electronic mail services, and cooperation and direct synchronization with other teleconferencing tools such as video, white boarding, and other collaborative software.

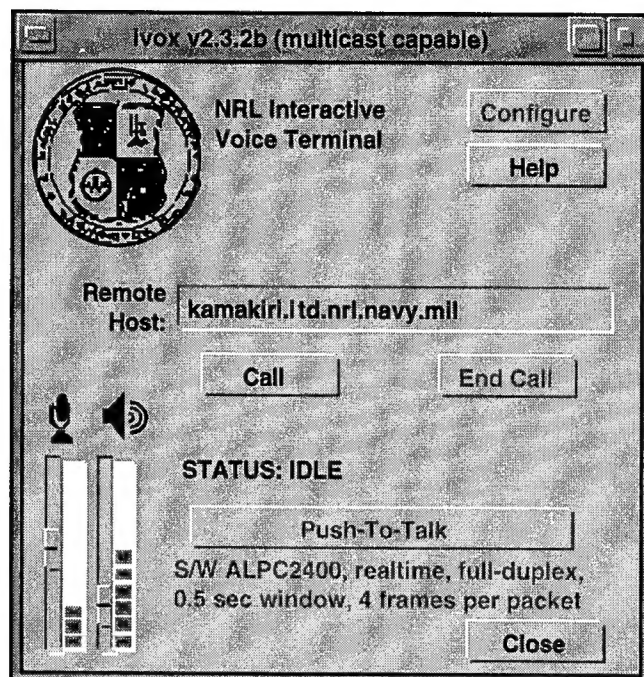


Fig. 1 - IVOX Graphical User Interface Main Window

Point-to-Point and Conference Calls

IVOX supports unicast (point-to-point) and multicast (conferencing) IP [12] communications. The current IVOX user interface was originally designed for point-to-point operation, and the capability for conferencing on an IP multicast address has recently been added. A future version of IVOX will include user interface additions that provide a listing of conference participants and indications of their activity.

IP multicast routing allows conferences with potentially hundreds (or more) of participants receiving network traffic while making efficient use of communication resources (i.e., network traffic is duplicated only when absolutely necessary). To create an IP multicast conference in IVOX, the participants need to agree on an IP multicast address (e.g., 224.x.x.x) in advance. The participants then join the conference by entering the IP multicast group address into the IVOX "Remote Host" text field and click on the "CALL" button. Participants may join and leave the conference any time at will. The host workstations and routers take care of the rest.

IVOX can also operate with the *Session Directory* (sd) application that is commonly used for advertising and establishing IP multicast conferences on the Internet multicast backbone (MBONE). IVOX may be launched by sd with command line options specifying the IP group address, packet time-to-live (ttl), and voice compression parameters for the conference session. Internet World Wide Web (WWW) servers and browsers can also be used to advertise and initiate IP multicast conferences. IVOX command line options can be used to allow web browsers to launch IVOX with the appropriate set of parameters. For wide area conferencing, routers in the path(s) between conference participants must support IP multicast forwarding and group management. Major commercial router vendors are now including support for IP multicast as a standard router feature.

Multiple Voice Compression Rates and Communication Modes

IVOX supports voice compression algorithms that operate at 2400, 1200, 800, and 600 bps. IVOX is also capable of operating with external voice compression hardware. For general purpose use, the 2400 bps algorithm is the best choice. This provides intelligible speech with modest throughput requirements. The lower throughput and somewhat lower quality algorithms are provided for situations where the minimum data rate possible is necessary. Additionally, IVOX provides a *non-real-time* mode to allow for limited interactive voice communication when the network is not capable of supporting real-time voice at any data rate. Future versions of IVOX will support higher data rate, higher quality voice coding techniques for operation on high throughput network connections.

Full-duplex and half-duplex communication modes are supported. With full-duplex operation, any party may speak at any time. A mode that enforces a half-duplex discipline on the users is provided for point-to-point operation. This half-duplex discipline facilitates productive conversations that may occur over long-delay network connections (e.g., satellite links).

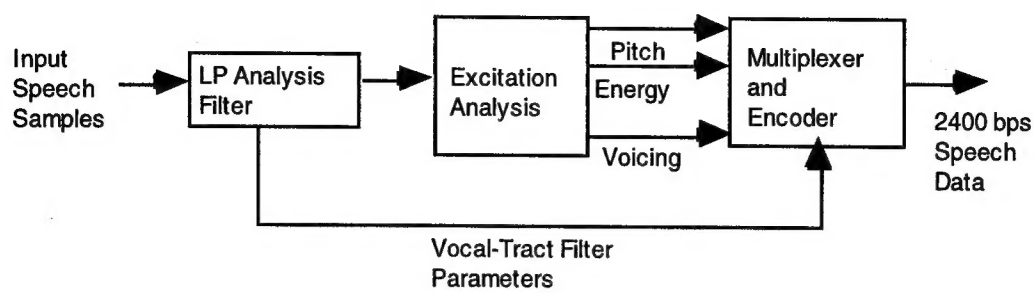
Voice Algorithm Issues

Use and Enhancement of FS-1015 LPC Vocoding

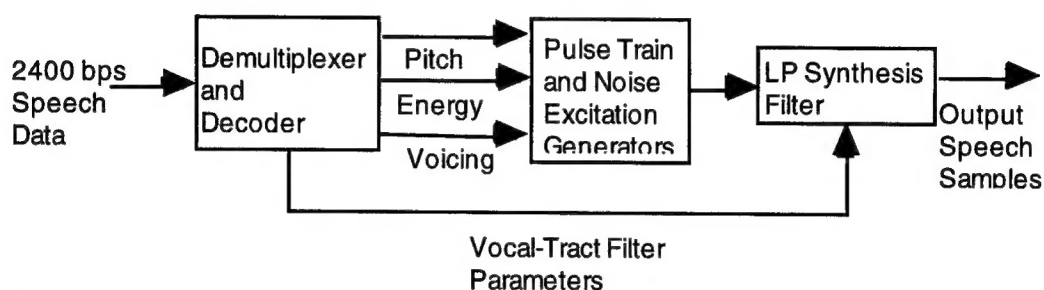
At present, IVOX makes use of a set of multiple rate vocoders based upon the FS-1015 2400 bps LPC-10 vocoding algorithm [18]. Figure 2 shows a block diagram of the LPC-10 vocoder. Linear prediction analysis is performed at a 22.5 ms frame rate by an open loop tenth-order covariance method. IVOX uses this algorithm directly for constant bit rate (CBR) service at 2400 bps. NRL Code 5520 has also developed a variable rate processing enhancement to the LPC-10 algorithm and created a library of LPC-based voice compression software routines for use in IVOX [14].

There are two important frame-by-frame estimation parameters, energy and voicing, which are used to perform variable rate encoding decisions. A frame-by-frame root mean square (rms) energy measurement is already implemented by the LPC analysis routine, and its output is used to feed the silence detection processor. In addition, unvoiced LPC frames contain redundant information that can be removed for network voice applications, thus resulting in a shorter frame length.

The FS 1015 LPC-10 standard uses a Hamming error correction coding method that improves performance in high bit error rate environments. As discussed earlier, within a network environment, the transport, network, and link layers generally provide delivery of error-free datagram data, although order and arrival times are not always guaranteed. Therefore, the Hamming error correction coding serves little purpose in the network voice application. We use this rationale to create a variable frame structure and allow silence frames to be dropped. If the energy value for an unvoiced frame is below a preset or adaptive threshold value, it is considered a silence frame and need not be transmitted. The resulting vocoder transmission source will produce vocoder frames during "voice spurt" periods in which the frame rms energy is above a fixed or adapted threshold. To make this work, the Hamming decoder within LPC-10 has been deactivated and silence frame processing has been added. Initial experiments and operational use with this variable rate vocoding scheme have demonstrated good intelligibility as well as efficient use of the available transmission bandwidth. Over one second interval measurements, typical voice conversations have indicated an average transmission rate requirement of approximately 1300 bps.



(a) LPC-10 Transmitter



(b) LPC-10 Receiver

Fig. 2: LPC-10 Voice Processor

Vector Quantization Compression Methods

To achieve voice digitization rates below 2400 bps, IVOX supports a number of vocoding algorithms based on vector quantization of raw FS-1015 LPC analysis output data. This vector quantization further compresses voice data and has been designed to introduce a minimum increase in overall signal distortion. Vector codebooks are being applied with a result in overall constant bit rate throughputs of approximately 600, 800, and 1200 bps. Figure 3 below describes the data flow.

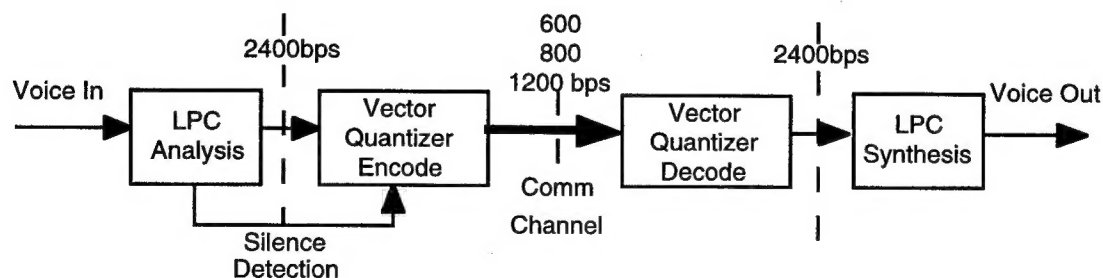


Fig. 3 — Vector quantization of LPC parameters

Fast tree search methods and hierarchical codebooks have been applied to the vector quantization coding to significantly reduce the processing speed requirements over brute force single codebook approaches. This has made it possible to achieve software-only operation for very large effective

codebook sizes (e.g., 25 bit or ~33 million entries) on standard workstations in real-time. Memory requirements have also been reduced by the use of parallel, hierarchical codebooks [15, 16].

Call Setup and Management Protocol

A robust call setup and session management protocol was designed to support IVOX. The protocol is based on concepts developed in the CSNI project [17]. The protocol was designed to be low overhead and provide reliable operation in the face of long communication delays and the potential for dropped packets in the tactical communication networks. The call setup protocol was enhanced for the NRL DVI ATD to work in conjunction with the resource reservation capabilities of that network. See Ref. 18 for further details.

Call setup for the calling party (caller) consists of the transmission of a CALL_REQUEST packet and subsequent receipt of a CALL_REPLY packet from the remote party (callee). The CALL_REPLY will be either a positive acknowledgment that the call is accepted or a rejection for one of several reasons (e.g., "busy" with another call, call rejected). The CALL_REQUEST packet is repeatedly transmitted until a valid CALL_REPLY is received or after a predetermined number of attempts have been made. Upon receiving a CALL_REQUEST, the IVOX callee is given the option of accepting or rejecting the call if they are not "busy" with another call. If the callee does not respond within a certain amount of time, the callee workstation will automatically reject the call. Calls are also automatically rejected if the local workstation does not support the vocoder type requested in the CALL_REQUEST packet.

The CALL_REQUEST Packet

Table 1 — CALL_REQUEST Packet Format

Field Name	#Bits	Possible Values
type	4	1 (0001)
priority	4	0-15 (Priority for CLNP)
metric	7	CLNP Routing Metric
comm_mode	1	0 = Half Duplex, 1 = Full Duplex
vocoder_type	(8*8)	Vocoder name (up to 8 characters, zero padded)
num_frames	8	Number of vocoder frames/ packet
session_id	8	xxxx0000 (Caller x fills in 4 msb's)
caller name	(8*8)	User name (up to 8 characters, zero padded)
callee name	(8*8)	User name (up to 8 characters, zero padded)

The CALL_REQUEST packet is identified by a type field that has a value of 1. The *priority* and *metric* fields in the packet were provided simply for CSNI to inform the remote user of the call priority and routing metric chosen for the session by the caller. The *comm_mode* (communications mode) field indicates if the session is to be conducted in a half duplex or full duplex fashion. The *vocoder_type* and *num_frames* (number of frames) fields indicate to the callee which vocoder algorithm is to be used for the call and the number of vocoder frames packaged into each packet. The *vocoder_type* field is a short string (up to 8) of ASCII characters. Some example vocoder type strings include "ALPC2400," "LPC2400," "ALVC800" "LVC600," "CELP4800," and "ULAW64K."

The *session_id* (session identification) field is a number synthesized at call setup time to uniquely identify the current call in place. This is used to filter packets from previous session that are delivered late. The reliable link layer protocols used in the CSNI project could result in packets from previous sessions interfering with a current session because of late delivery. The session identifier is used in conjunction with the source address of packets during the voice session to determine their validity to the ongoing session. The *session_id* value is negotiated at call setup with the caller filling the four most significant bits in the *CALL_REQUEST* and the callee filling in the four least significant bits in the subsequent *CALL_REPLY* message. The filler bits are randomly selected by each IVOX application at startup and incremented over the course of call attempts so that unique session identifiers are probably for call sessions close in time.

The *caller_name* and *callee_name* fields allow IVOX to display the login name of the user placing the call. It allows the IVOX terminal who is called to display for which local user the call is intended. These fields are unused in more recent versions of IVOX because computer hosts generally have a single audio device at the console, and as a result all IVOX calls are directed to the user who “owns” the audio device.

CALL_REPLY Packet

The *CALL_REPLY* packet is sent back to the calling IVOX terminal when the user accepts or rejects the call via the graphical user interface, or when a time-out condition occurs. The value of the *type* field is equal to 2 for the *CALL_REPLY* packet and the format of this IVOX packet type is described in Table 2.

Table 2 — *CALL_REPLY* Packet Format

Field Name	#Bits	Possible Values
<i>type</i>	4	2 (0010)
<i>reason</i>	4	<i>CALL_ACCEPTED</i> , <i>CALL_REFUSED</i> , <i>CALLEE_BUSY</i> , <i>CALLEE_UNAVAILABLE</i> , <i>CALLEE_UNKNOWN</i> , <i>VOCODER_UNSUPPORTED</i> , <i>BAD_SESSION_ID</i>
<i>session_id</i>	8	xxxxyyyy (Callee y fills in 4 lsb's)

The *reason* field is used to indicate if the call is accepted or the reason why it was rejected. Table 3 lists the current set of possible values for the *reason* field.

VOICE_DATA Packet

Once a call session has been established, voice communication is accomplished with the transmission of *VOICE_DATA* packets. Table 4 provides an overview of the format of the *VOICE_DATA* packet type. The *type* field has a value of 3 for the *VOICE_DATA* packet type.

Each *VOICE_DATA* packet has a *seq_number* (sequence number) field. The value of the sequence number determines the packet's timing in relation to other voice packets. This allows for the IVOX to properly reorder any packets that arrive out of order and to appropriately time voice playback in the face of missing (dropped by the network or silent) voice packets. The *session_id*

Table 3 — CALL_REPLY Reason Values

Value	Reason
0	CALL_ACCEPTED
1	CALL_REFUSED (by user)
2	CALLEE_BUSY
3	CALLEE_UNAVAILABLE
4	CALLEE_UNKOWN
5	VOCODER_UNSUPPORTED
6	BAD_SESSION_ID

Table 4 — VOICE_DATA Packet Format

Field Name	#Bits	Possible Values
type	4	3 (0011)
seq_number	12	0-4095
session_id	8	xxxxyyyy (caller-callee tuple = 0-255)
ptt_id	8	0-255
voice_data	8*n	n depends on the vocoder type and number of frames per packet.

field contains the value negotiated at call setup. The *ptt_id* field uniquely identifies the voice data as part of the a voice sequence occurring under the same instance of enabling IVOX push-to-talk (PTT) control. This field is used to filter very late-arriving packets. It is most useful for IVOX's non-real-time voice communication mode where voice from a single press of the PTT button (for up to 30 s) is buffered for playback until the *entire* voice segment is received. This allows for limited but interactive voice conversations on network connections that have insufficient throughput to support real-time voice conversations. The *voice_data* field contains the encoded voice data. The length of this field depends on the particular vocoder type and number of vocoder frames in each packet. For example, the ALPC2400 vocoder algorithm with four frames per packet results in 27 bytes in the *voice_data* field.

TOKEN_RELEASE Packet

When the user releases the PTT control, a TOKEN_RELEASE packet is transmitted. This lets the remote IVOX terminal know that the local user is finished speaking. This is used to enforce the half-duplex mode of operation and is also useful for the IVOX software during multicast conference operation where many users are taking turns in speaking. The type field value is equal to 4 for the TOKEN_RELEASE packet. Table 5 describes the format of the TOKEN_RELEASE packet.

The *seq_number* field serves the same purpose as it does for voice data. It provides timing information so that IVOX can properly time completion of playback of the current voice segment. The *session_id* and *ptt_id* fields are the same as those for voice data so that a TOKEN_RELEASE packet received may be used with the correct set of voice data.

Table 5 — TOKEN_RELEASE Packet Format

Field Name	#Bits	Possible Values
type	4	4
seq_number	12	0-4095
session_id	8	xxxxyyyy (caller-callee tuple = 0-255)
ptt_id	8	0-255

CALL_END Packet

When either party, caller or callee, wishes to terminate a call session, the **CALL_END** packet is transmitted. The **CALL_END** packet is used to let the IVOX terminal know that it is time to return to an idle state and wait for another call. The **CALL_END** packet has a *type* field value equal to five and its format is described in Table 6.

Table 6 — CALL_END Packet Format

Field Name	#Bits	Possible Values
type	4	5
seq_number	12	0-4095
session_id	8	xxxxyyyy (caller-callee tuple = 0-255)
ptt_id	8	0-255

The *seq_number* and *ptt_id* fields are used to properly time the ending of the call in relation to received voice data. The *session_id* is used to properly identify the session to be ended. A **CALL_END** message with a valid *session_id* is never ignored even if the *seq_number* and *ptt_id* are out of bounds in relation to the IVOX receiver's current timing.

Voice Packet Buffering and Reordering

Connectionless network packet delivery can result in variances in packet arrival times relative to the original transmit timing, and packets are sometimes delivered out of order. IVOX provides a buffering and resequencing mechanism to recover the voice packets' original ordering and transmit timing. IVOX maintains a sliding buffer "window" in which reorders packets for playback. The buffer window introduces end-to-end delay for voice communication, so IVOX adaptively adjusts the size of this buffer to minimize delay while maintaining good voice quality. The size of the window is dependent upon network performance. Network performance can vary over time, as other loading is placed on the network and depending upon the data delivery characteristics of the underlying communication media. IVOX maintains statistics on the arrival times of voice packets with respect to the expected arrival time given the voice coding technique and voice packet content. These statistics are used to adjust the IVOX window buffer size and offset in relation to received voice data packets.

IVOX uses a rule-based mechanism for adjusting the buffer size. For example, if IVOX begins losing a large number of packets because its buffer time is too short, it will quickly adjust to recover voice quality. If IVOX measures long periods of good performance, it slowly reduces its buffering time to reduce overall end-to-end voice delay. IVOX will make adjustments during silent portions of speech to reduce the impact on speech quality. This same buffering mechanism that compensates for

network delay variations also helps compensate for workstation execution time variations as the workstation CPU demand fluctuates over time.

SUMMARY

Accomplishments

Since its conception and development, IVOX has become an operational fleet software component and has played a key role in numerous research projects and demonstrations. IVOX was adopted as a core feature of the recent Joint Deployed Intelligence Support System (JDISS) software release. IVOX was used to demonstrate low-bandwidth network voice capability from an operational NRL booth during the 1995 Armed Forces Communications and Electronics Association (AFCEA 95) convention.

Additionally, IVOX has been integrated into Tactical Aircraft Mission Planning System (TAMPS) and Common Operation Mission Planning and Support Strategy (COMPASS) workstation terminals for the 1995 Joint Warrior Interoperability Demonstration (JWID '95) demonstration. During JWID '95, IVOX generally provided good results with occasional degraded service being noted during periods of high network loading. This result was expected and future use of resource reservation techniques will allow IVOX to provide good voice communications even during these periods of high network loading. IVOX enhancements that included some unique resource reservation capabilities were successfully developed and demonstrated during the Phase 2 NRL DVI ATD field test on the Chesapeake Bay in the summer of 1995. IVOX continues to play a role in the CSNI project and has been demonstrated successfully in both real-time and non-real-time modes.

Future Directions

With the increasing proliferation of distributed collaborative tools, awareness data, and integrated C4I networking for the warfighter, real-time network voice applications will become a software component of increasing value. IVOX is presently providing an early capability in this area with little or no additional system cost.

Future efforts to enhance IVOX include higher data rate vocoding (e.g., 4800 bps Code Excited Linear Prediction (CELP)) for improved speech quality and support for emerging resource reservation setup protocols, such as the Resource Reservation Protocol (RSVP). RSVP will provide a standard method for applications to attain a guaranteed quality-of-service in a shared network environment. It is anticipated that the importance of bandwidth efficient, voice conferencing using multicast networking technology will increase. Improved features for multicast conferencing are planned for IVOX.

IVOX will provide a useful starting point for exploring options in session establishment and management for distributed conferencing and mission planning and coordination. IVOX has the potential to be closely coupled with other distributed communication applications to obtain robust, adaptive data rate operation where applicable.

ACKNOWLEDGMENTS

We are grateful to the Office of Naval Research (ONR) for sponsoring and supporting the early work on IVOX under a number of projects. Tom Tremain, Mary Kohler, and John Collura of NSA deserve special thanks for their technical assistance and excellent work on very low data rate voice

digitization algorithms and real-time implementations. Special thanks also go to Don MacLachlan of Canada's Communications Research Centre (CRC) for his assistance in a number of programming issues during the development of IVOX, and also to the rest of the CSNI Task 2B working group for their contributions in the development of the protocols employed by IVOX. We also acknowledge the insight of the JDISS program office and the TAMPs program office (including George Kamis at NRL) for recognizing the value of IVOX for operational communication and for encouraging and supporting the early transition of IVOX for Navy use.

REFERENCES

1. R.B. Adamson, "IVOX, the Interactive VOice eXchange Application," NEWLINK Global Engineering Tech Rep. 101, Sept. 1995.
2. W. Richard Stevens, "TCP/IP Illustrated," Vol. 1. (Addison-Wesley Publ. Co., Reading, MA, 1994).
3. Communications Support System Requirements Document, U.S. Navy Space and Naval Warfare (SPAWAR) Systems Command, 1995.
4. Henning Schulzrinne, "Voice Communication Across the Internet: A Network Voice Terminal," University of Massachusetts, July 26 1992.
5. R.B. Adamson, "Communication Systems Network Interoperability," 1995 NRL Review, pp. 146-149.
6. E.L. Althouse, J.P. Macker, J.P. Hauser, and D.J. Baker, "Integrated Services in Tactical Communication Systems," 1995 NRL Review, pp. 141-143.
7. "Data and Voice Integration Advanced Technology Demonstration (ATD): Execution Plan," submitted to ONR, 16 June 1992.
8. J.P. Macker and R.B. Adamson, "Proposed Draft Application Programmers Interface (API) for the Interactive VOice eXchange (IVOX) Vocoder Processes," NRL Tech. Mem. Rep., 5520-59, April 1994.
9. J.P. Macker and R.B. Adamson, "A Variable Rate Voice Coder using LPC-10E," NRL Tech. Mem. Rep., 5520-92, July 1994.
10. S. Floyd, "Link-Sharing and Resource Management Models for Packet Networks," Lawrence Berkeley Laboratory, Dec. 1994.
11. L. Zhang, R. Braden, et al, "Resource ReSerVation Protocol (RSVP) - Version 1 Functional Specification," Internet Draft, Nov. 1994.
12. S.E. Deering and D.R. Cheriton, "Host Groups: A Multicast Extension to the Internet Protocol," Internet RFC 966, Dec. 1985.
13. Federal Standard 1015, "Analog to Digital Conversion of Voice by 2400 bits/s Linear Predictive Coding," published by GSA, Nov. 1984.

14. J.P. Macker and R.B. Adamson, "Linear Predictive Coding (LPC) Speech 'C' Software Library," NRL Tech. Mem. Rep., 5520-65A, April 1994.
15. Mary A. Kohler, "Real Time Low Rate Voice Coding," Version 1.0, NSA Tech. Rep., TR-R22-10-93, May 1993.
16. J. S. Collura and T.E. Tremain, "Vector Quantizer Design for the Coding of LSF Parameters," Draft NSA Tech. Rep., 1993.
17. B. Awada and R.B. Adamson, "CSNI Interactive Packet Voice Scheme," CSNI Document TASK2B/CA/003, 10 Nov. 1993.
18. J.P. Macker, "Real-Time Voice Delivery Support for Integrated Mobile Military Internetworking," NRL Rep. in Review, Nov. 1995.